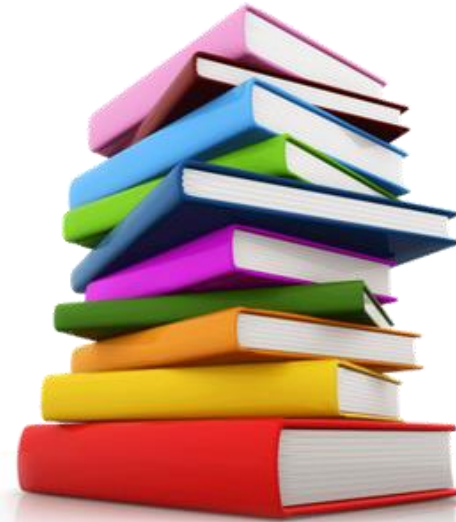


Cluster Analysis




Data Files:

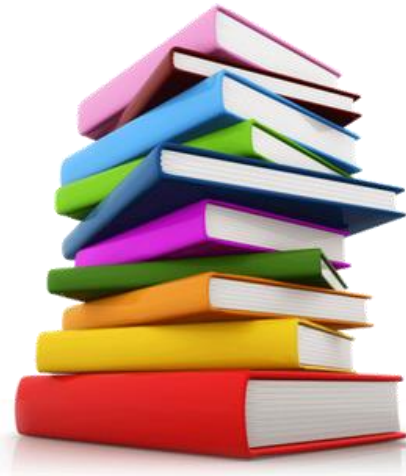
1. How To Stop Worrying & Start Living
2. The Future of Football
3. Ian Fleming Author Data Sheet



Data Sets

> This PC > Desktop > clusterTM

Name ^	Date modified	Type	Size
 bond	19-08-2020 21:36	Text Document	4 KB
 dc	18-08-2020 19:22	Text Document	14 KB
 fb	19-08-2020 01:37	Text Document	15 KB



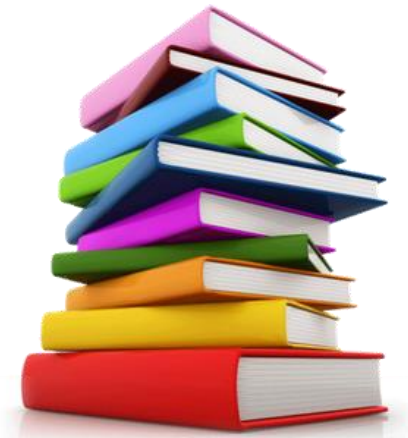
Ian Fleming Author Data Sheet by Macmillan



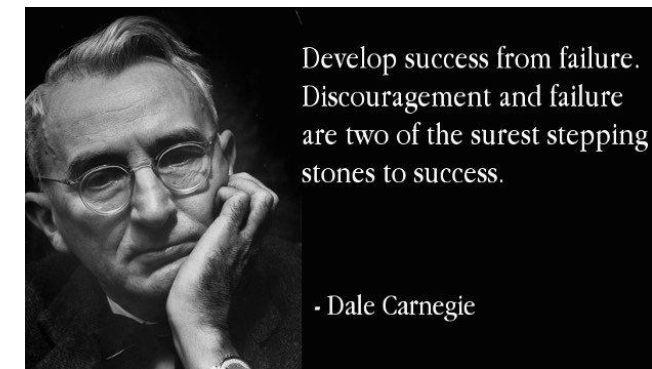
Dr Vinod on Cluster in Text Data 8971073111
vinodanalytics@gmail.com


```
1 # Jesus is my Saviour!
2 setwd("C:/Users/Dr Vinod/Desktop/clusterTM")
3 getwd()
4 library(tm)
5 #Create Corpus
6 docs <- Corpus(DirSource("C:/Users/Dr Vinod/Desktop/clusterTM"))
7 docs # 3 documents importes
8 length(docs) # 3
9
```

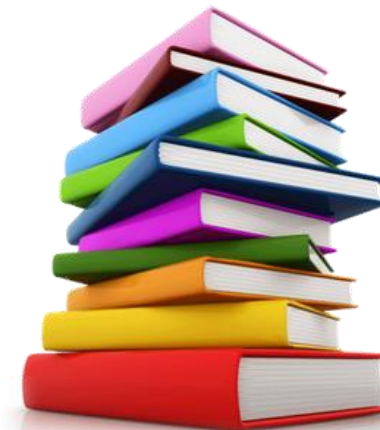
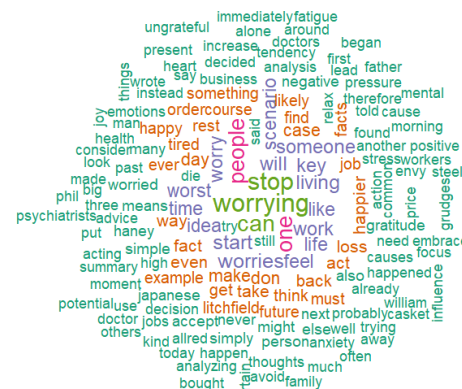
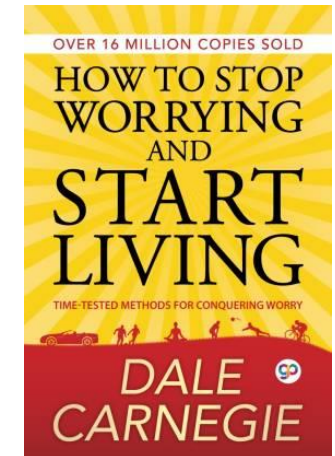
```
> dtm # sparsity 61%, 2276/3511, Max term length 18, tf
<<DocumentTermMatrix (documents: 3, terms: 1929)>>
Non-/sparse entries: 2276/3511
Sparsity           : 61%
Maximal term length: 18
Weighting          : term frequency (tf)
> docs # 3 documents importes
<<SimpleCorpus>>
Metadata: corpus specific: 1, document level (indexed): 0
Content:  documents: 3
> length(docs) # 3
[1] 3
```




```
12
13 writeLines(as.character(docs[[2]])) # dc
14
```



```
> writeLines(as.character(docs[[2]])) # dc
i stop worrying start living summary review dale carnegie stop worrying start living dale car
negie sitting reading list pick key ideas book quick summary everyone worries us realize thou
gh damaging health worrying excessively can actually make physically ill plato already knew m
ind body intimately linked fact mayo brothers famous physicians declared half hospital beds o
ccupied people suffering frustration anxiety worry despair arthritis example one many debilit
ating conditions can brought worry fact two leading causes arthritis worry related marital sh
ipwrecks financial woes also medical cases suggest worrying can increase likelihood insanity
diabetes clearly worrying bad health unfortunately us causes worry everywhere they re especia
lly common work high pressure jobs tend generate worrying predictably illness calmer tranquil
jobs high stress levels associate high pressure jobs can lead heart disease one study showed
third business executives suffer heart disease stomach ulcers high blood pressure another st
udy found twenty times many doctors farm workers die heart failure means you re highly stress
```




```

14
15 writeLines(as.character(docs[[3]])) # fb
16

```

```

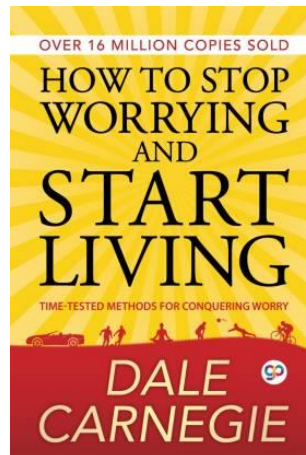
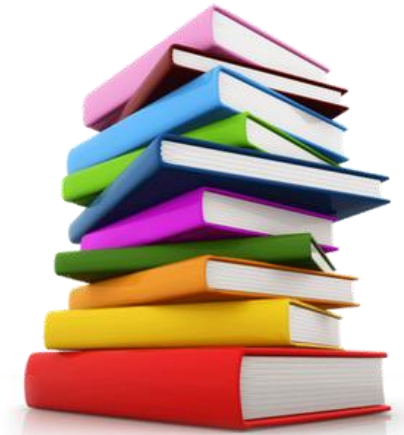
> writeLines(as.character(docs[[3]])) # fb
see discussions stats author profiles publication https www researchgate net publication futu
re football article society august doi citation reads author william beaver robert morris uni
versity publications citations see profile authors publication also working related projects
campus sexual assault view project content following page uploaded william beaver september u
ser requested enhancement downloaded file soc doi commentary future football william beaver p
ublished online may springer science business media new york three popular spectator sports w
ill decline sports happens united states baseball boxing horse racing will due scandals exces
sive violence began change growing popularity football fact research increasingly shows recei

```



```
17 dtm <- DocumentTermMatrix(docs)
18 dtm # sparsity 61%, 2276/3511, Max term length 18, tf
19 #inspect(dtm)
20
```

```
> dtm <- DocumentTermMatrix(docs)
> dtm # sparsity 61%, 2276/3511, Max term length 18, tf
<<DocumentTermMatrix (documents: 3, terms: 1929)>>
Non-/sparse entries: 2276/3511
Sparsity           : 61%
Maximal term length: 18
Weighting           : term frequency (tf)
```



```

19
20 #inspect(dtm)
21 inspect(dtm[1:3,100:105])
22

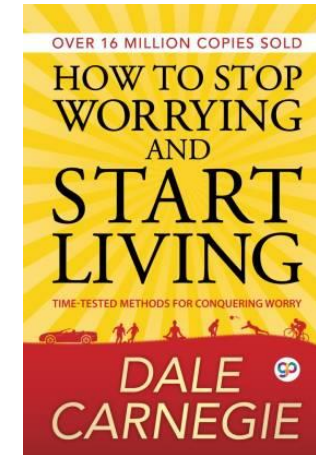
```

```

> inspect(dtm[1:3,100:105])
<<DocumentTermMatrix (documents: 3, terms: 6)>>
Non-/sparse entries: 8/10
Sparsity           : 56%
Maximal term length: 10
Weighting          : term frequency (tf)
Sample            :

```

Docs	golden	goldfinger	gone	got	groups	gun
bond.txt	1	1	1	1	2	1
dc.txt	0	0	0	2	0	0
fb.txt	0	0	0	1	0	0

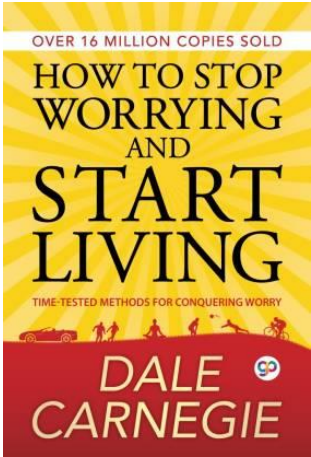



```
23
24 m <- as.matrix(dtm)
25 write.csv(m,file="3docs2.csv")
26
```



This PC > Desktop > clusterTM

	Name ^	Date modified	Type	Size
✦	3docs2	20-08-2020 01:29	Microsoft Excel Co...	30 KB
✦	bond	19-08-2020 21:36	Text Document	4 KB
✦	dc	18-08-2020 19:22	Text Document	14 KB
✦	fb	19-08-2020 01:37	Text Document	15 KB

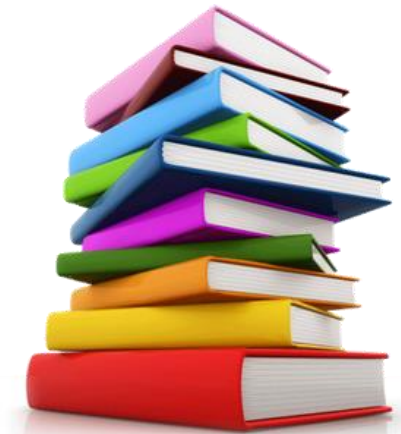
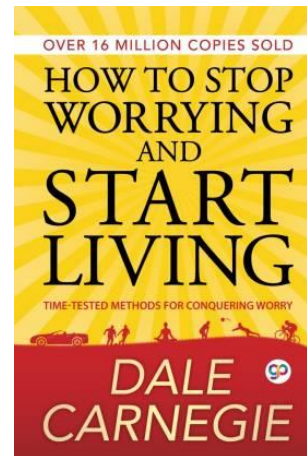


	A	B	C	D	E	F	G	H	I	J	K	L
1		actually	add	aged	agent	agents	also	always	ann	anne	another	archive
2	bond.txt	1	1	1	2	2	1	1	2	1	1	1
3	dc.txt	1	0	0	0	0	5	2	0	0	4	0
4	fb.txt	0	0	1	0	0	9	7	0	0	0	0

```
26  
27 distMatrix <- dist(m, method="euclidean")  
28 distMatrix  
29
```



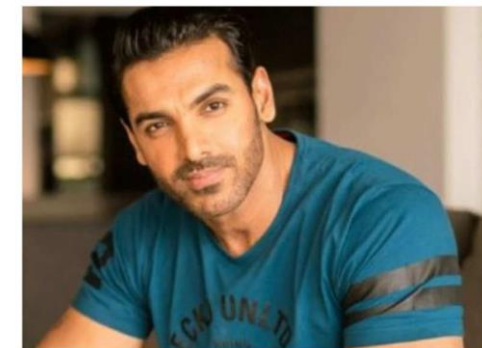
```
> distMatrix  
      bond.txt  dc.txt  
dc.txt 104.1489  
fb.txt 111.2385 128.4484
```



Euclidean Distance

Euclidean Distance					
	HND	ACT	SB		
AK	8	9	8		
SK	7	7	7		
JA	9	6	7		
AK-SK	=SQRT((8-7)^2+(9-7)^2+(8-7)^2)				2.44949
AK-JA	=SQRT((8-9)^2+(9-6)^2+(8-7)^2)				3.316625
SK-JA	=SQRT((7-9)^2+(7-6)^2+(7-7)^2)				2.236068

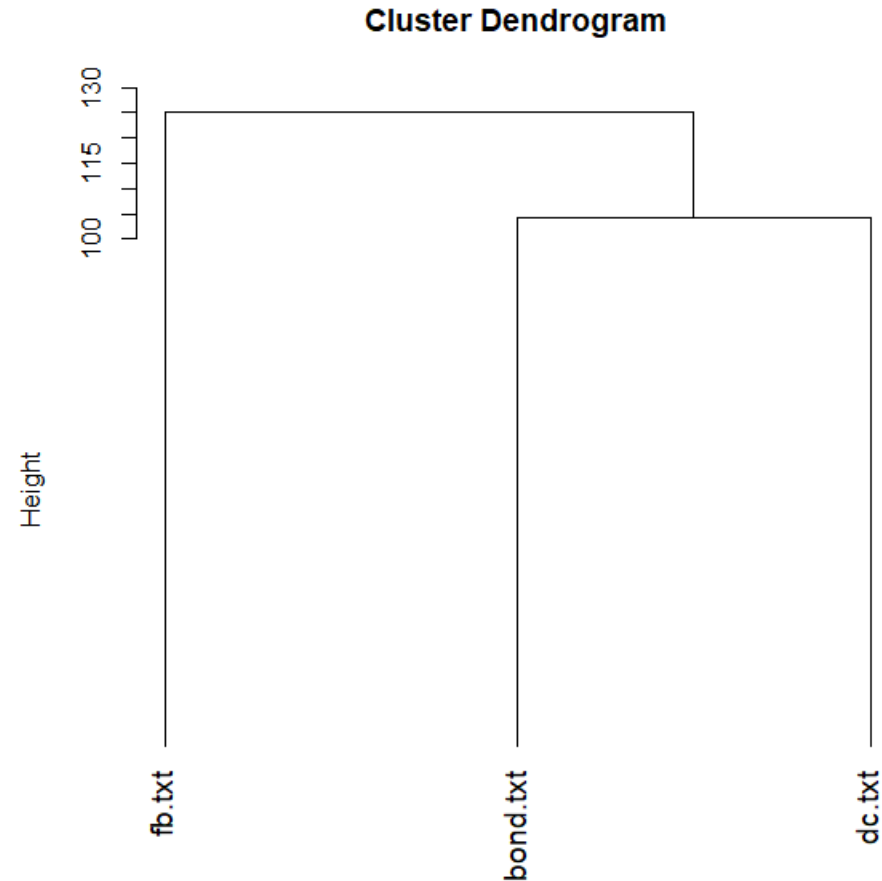
```
> distMatrix
      bond.txt  dc.txt
dc.txt 104.1489
fb.txt 111.2385 128.4484
```



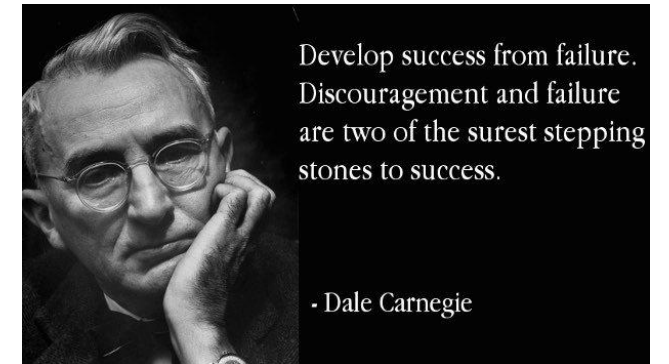
```

29
30 groups <- hclust(distMatrix,method="ward.D")
31 plot(groups, cex=1.2, hang=-1) # cex will change the font size
32

```



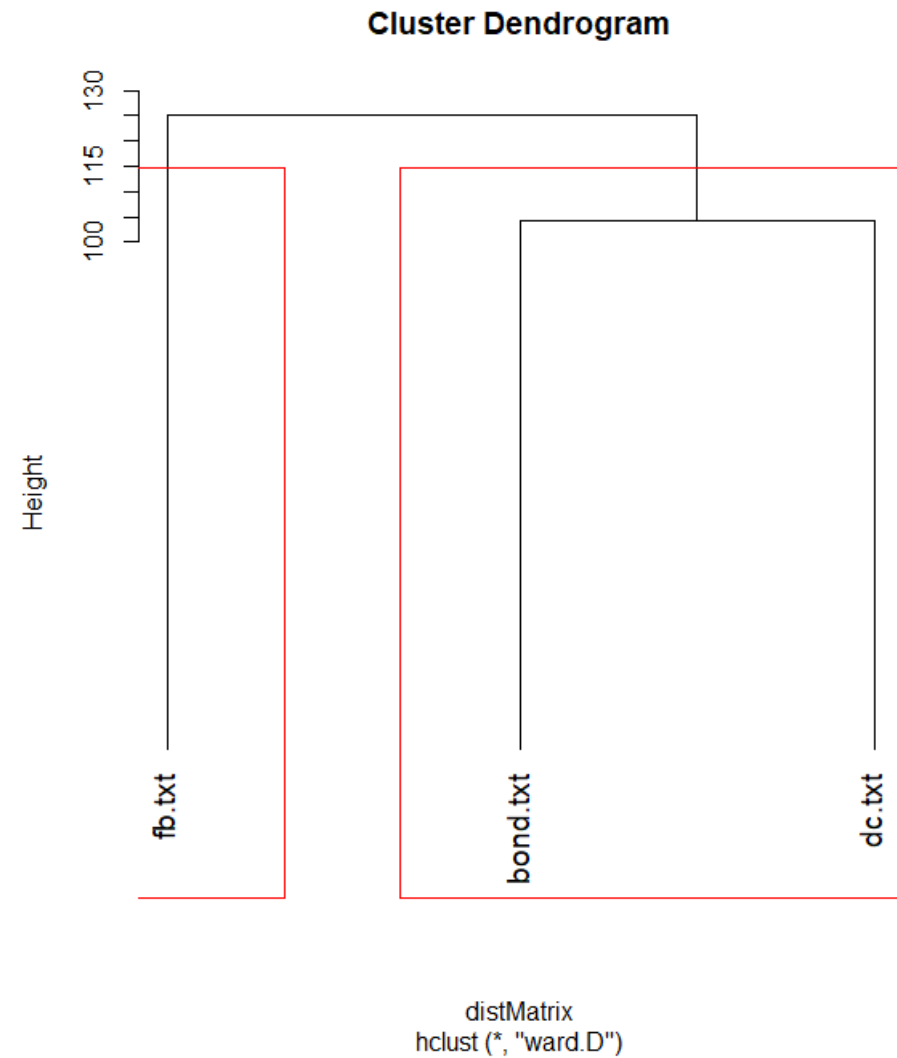
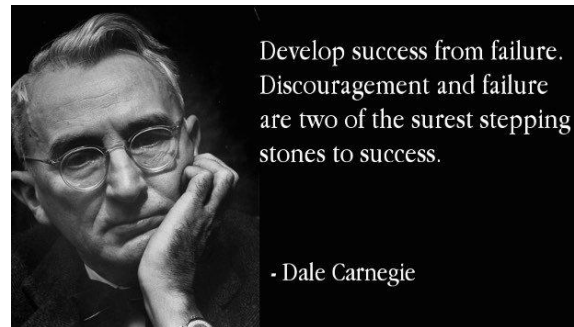
distMatrix
hclust (*, "ward.D")



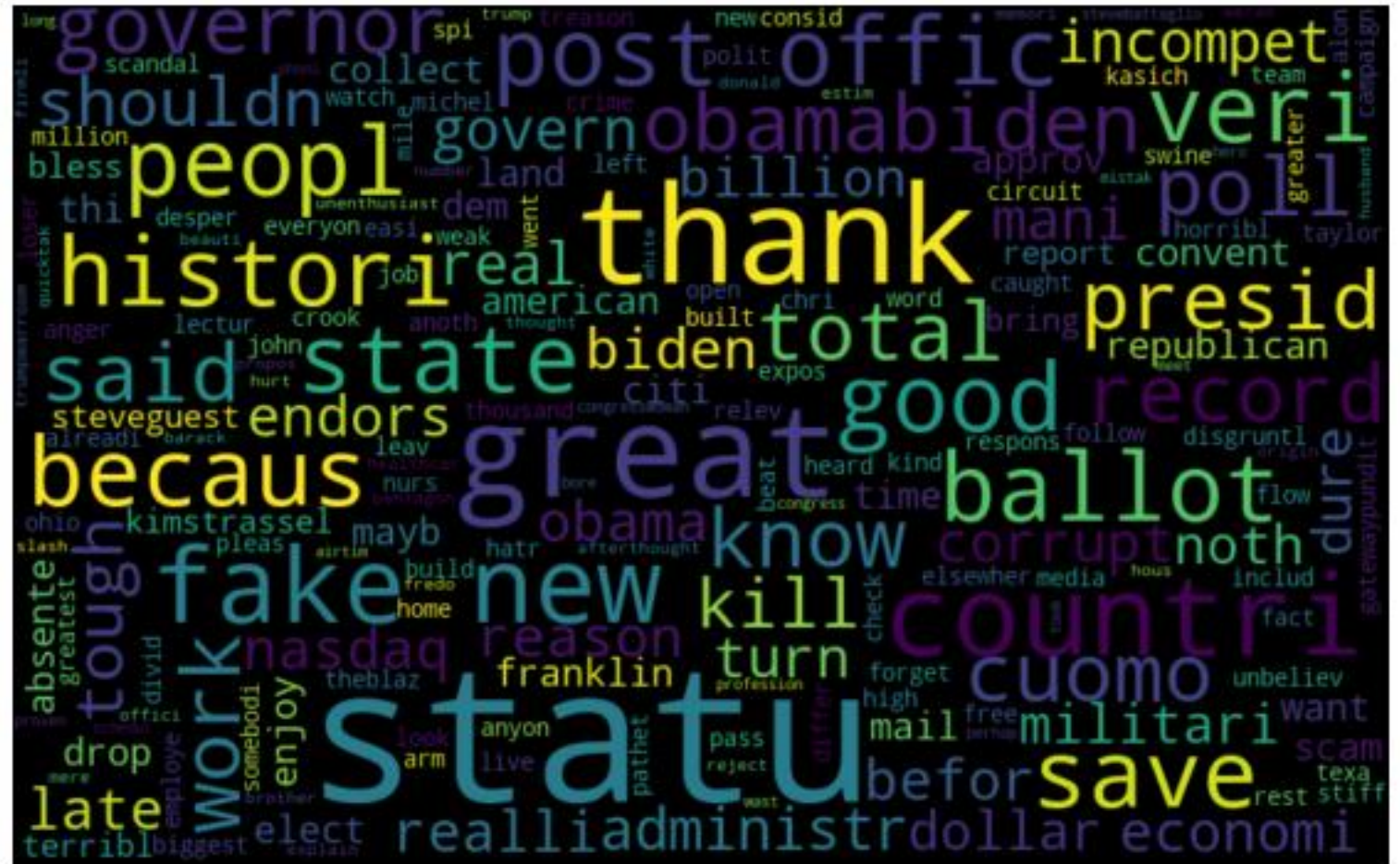
```

32
33 rect.hclust(groups, k=2)
34

```



19 March 2020 till 18th August 2020



Inbox (34,244) - vinodanalytics@ x | Inbox (2,012) - vinod.murti@ioag x | Remove matched patterns in a st x | (24) Ratan N. Tata (@RNTata2000) x +

twitter.com/RNTata2000

Apps Gmail YouTube Maps

Home Explore Notifications Messages Bookmarks Lists Profile More

Ratan N. Tata 165 Tweets

Ratan N. Tata @RNTata2000
Chairman Emeritus, Tata Sons. Chairman, Tata Trusts
Mumbai Joined April 2011
35 Following 9.1M Followers
Followed by Nandan Nilekani, Tuhin Chattopadhyay, and 3 others you follow

Tweets Tweets & replies Media Likes

Ratan N. Tata Retweeted

Tata Trusts @tatatrusters · Jun 19
Celebrating the spirit of self-reliance, we are proud of our weavers who, with the help of Antaran honed their entrepreneurial skills to support their families during the national lockdown. On behalf of these artisans, their communities and Antaran, we thank you for your support.

download.png twitter_bpcl.py

Search Twitter

You might like

MadhyaPradeshTo... @MPTourism Follow
Promoted

Amit Shah @AmitShah Follow

Amitabh Bachchan @SrBachchan Follow

Show more

What's happening

Show all

Type here to search

02:03 19-08-2020

DO
WHAT
YOU
LOVE
LOVE
WHAT
YOU
DO

“Choose a job
you **love**, and
you **will** never
have to work
a day in your
life. ”

Confucius

